Compressing high-resolution data through latent representation encoding for downscaling large-scale AI weather forecast model

Qian Liu[2], Bing Gong[3,1], Xiaoran Zhuang[4], Xiaohui Zhong[1], Zhiming Kang[5], and Hao Li[1,2]

[1]Artificial Intelligence Innovation and Incubation Institute, Fudan University, Shanghai, 200433, China
[2]Shanghai Academy of Artificial Intelligence for Science, Shanghai, 200232, China
[3]Department of Electrical and Computer Engineering Shanghai Normal University, Shanghai, China
[4]Jiangsu Meteorological Observatory, Nanjing, Jiangsu, 21008, China
[5]Nanjing Meteorological Observatory, Nanjing, Jiangsu, 21009, China

Correspondence: Zhiming Kang (angzm@cma.gov.cn) and Hao Li (lihao_lh@fudan.edu.cn)

Abstract. The rapid advancement of artificial intelligence (AI) in weather research has been driven by the ability to learn from large, high-dimensional datasets. However, this progress also poses significant challenges, particularly regarding the substantial costs associated with processing extensive data and the limitations of computational resources. Inspired by the Neural Image Compression (NIC) task in computer vision, this study seeks to compress weather data to address these challenges and en-

5 hance the efficiency of downstream applications. Specifically, we propose a variational autoencoder (VAE) framework tailored for compressing high-resolution datasets, specifically the High Resolution China Meteorological Administration Land Data Assimilation System (HRCLDAS) with a spatial resolution of 1 km. Our framework successfully reduced the storage size of 3 years of HRCLDAS data from 8.61 TB to just 204 GB, while preserving essential information. In addition, we demonstrated the utility of the compressed data through a downscaling task, where the model trained on the compressed dataset achieved

10 accuracy comparable to that of the model trained on the original data. These results highlight the effectiveness and potential of the compressed data for future weather research.

1 Introduction

Weather forecasting is crucial for society and various industries, supporting informed decision-making in areas such as agri-

15 culture, transportation, disaster management. Traditionally, numerical weather prediction (NWP)) models have been employed for this purpose, but they are computationally intensive, requiring substantial computing resources. Recent advancements in deep learning offer promising alternatives to NWP models, potentially offering faster and equally accurate forecasts (Bi et al., 2023; Li et al., 2023; Lam et al., 2022). However, the effectiveness of deep learning applications in weather and climate relies heavily on the availability of large-scale datasets. The computational demands for acquiring, storing, and managing such data

20 frequently exceed the capabilities of researchers with only modest setups, creating significant barriers for those lacking access to high-performance computing resources and data storage.

Moreover, to ensure accurate weather forecasting, numerous super-computing and research centers around the world conduct operational weather and climate simulations multiple times daily. For example, the European Centre for Medium-Range Weather Forecasts (ECMWF) manages 230 petabytes (PB) of data and processes approximately 600 million Earth observations

25 each day. This data volume is projected to quadruple over the next decade due to increasing spatial resolution in forecasting models (Klöwer et al., 2021). While this data growth provides more opportunities for training deep learning models, it also poses challenges as the sheer data volume can overwhelm the existing super-computing infrastructure and complicate the distribution of weather products due to limited network bandwidth. Thus, effective data compression techniques are essential.

Compressing weather data is similar to the Neural Image Compression (NIC) task in the computer vision domain, where

30 the goal is to reduce the file size of images while maintaining acceptable quality. For example, Ballé et al. (2016) employed an end-to-end convolutional neural network for image compression, while Chen et al. (2022) developed a framework to compress automotive stereo videos by reducing temporal redundancy. Most of the prior work has focused on using Variational Autoencoders (VAEs) for image compression (Ballé et al., 2018; Cheng et al., 2020). For instance, Cheng et al. (2020) demonstrated that employing discredited Gaussian Mixture Likelihoods to parameterize the latent code distributions achieves greater

35 accuracy than traditional entropy models.

Although NIC based on VAEs shares similarities with weather data compression in exploring latent representation patterns, significant differences remain. Natural images, with high correlation among red, green, and blue (RGB) channels, allow traditional compression techniques to exploit these relationships effectively. In contrast, while weather data exhibits correlations among its variables, these relationships are not as straightforward correlated as nature images. Thus this further complicates the

40 application of conventional methods. Additionally, the volume of weather data is substantially large. For example, the ERA5 reanalysis dataset Hersbach et al. (2020) consists of hourly data at a spatial resolution of 31 km, with 640×1280 points, totaling approximately 226 terabytes (TB))]. In comparison, the High Resolution China Meteorological Administration Land Data Assimilation System (HRCLDAS) Shi et al. (2014); Han et al. (2020a) offers a regional product with a spatial resolution of 1 km, incorporating 4500×7000 grid points hourly. Managing such large datasets for weather applications introduces chal-

45 lenges for NIC in deep learning, including increased data loading times and processing requirements that can hinder training and model convergence, ultimately degrading compression performance. Thus, this study proposes a new data compression method tailored specifically to weather data to address these challenges.

Furthermore, to evaluate the effectiveness of the proposed data compression method, we employed statistical downscaling as a proof-of-concept downstream application. Despite advancements in deep learning for downscaling, the conflict between the

50 need for extensive data and limited computational resources remains a significant challenge. For instance, Leinonen et al. (2020) utilized a generative adversarial model to stochastically downscale coarse radar data to high resolution without additional conditional variables. Harris et al. (2022) divided the entire UK region into smaller patches to incorporate more conditional inputs given computational constraints. yet this approach limits the deep learning model's ability to capture global features. Similarly, Zhong et al. (2024b) used high-resolution assimilation data as ground truth for downscaling, but their application

55 was confined to specific small regions in China. Thus, demonstrating the successful application of compressed data for these tasks could serve as a promising example in the field, alleviating computational constraints and promoting broader application of deep learning techniques in weather downscaling.

In addressing the challenges of managing large-scale weather data, we propose a NIC framework specifically designed for this purpose, compressing high-resolution weather data into a latent representation space to ease the data burden for down-60 stream tasks. The main contributions of this research are as follows:

– **Introduction of a novel data compression framework tailor to weather data:** We present a Variational Autoencoder (VAE) method for reducing high-resolution weather data. The VAE's encoder generates a quantize latent Gaussian distribution through the variational inference process. Our comparison of training strategies reveals that pre-training followed by a fine-tuning yields the best reconstruction performance, resulting in superior latent representation of the original 65 data.

– **Proof of Concept with large HRCLDAS Data**: We applied our framework to 1-km high-resolution HRCLDAS data, successfully compressing three years' worth of data from a total of **8.61 TB** to a compact **204 GB** data. Reconstruction results demonstrate that the VAE decoder can accurately reproduce the raw data with minimal loss of information and detail, effectively recovering and preserving key properties of the original data, such as extreme values.

70 – **Downstream Application - statistical downscaling:** We provide a downstream application - statistical downscaling the output from a deep learning based weather forecasting model, i.e. FuXi Chen et al. (2023), over China as proof of concept. Unlike most studies that resolve the NWP outputs, our research is the first to downscale deep learning-based weather forecasting model outputs, enhancing their resolution. Our results indicate that our framework significantly reduces computational costs while maintaining model performance.

75 ## 2 Methodology

2.1 Data sources

To assess the performance of the proposed compression framework, we employed downscaling as a proof-of-concept application, using HRCLDAS as the high-resolution ground truth. HRCLDAS Shi et al. (2014); Han et al. (2020a) is a blended dataset that integrates station observations, satellite data, and NWP data station advanced land surface and data assimilation techniques 80 (Han et al., 2020b; Zhong et al., 2024b). In this study, the HRCLDAS data encompasses three years of data from 2019 to 2021, including hourly 2-meter temperature (T_{2M}), 10-meter u-component of wind (U_{10M}), and 10-meter v-component of wind (V_{10M}). Its original dimensions are $4384 \times 6880 \times 3$, covering latitudes from $15°$ to $55°$ and longitudes from $75°$ to $135°$. We used the data from May 2019 to October 2019 and July 2020 to August 2021 as training dataset, and from September 2021 as testing for data compression task.

85 For the downscaling task, we used forecasts from the FuXi-2.0 model as low-resolution inputs. The FuXi-2.0 model is a cascaded machine learning weather forecasting model that provides 15-day global forecasts with a temporal resolution of 1

hour and a spatial resolution of 0.25 °Chen et al. (2023); Zhong et al. (2024a). In this study, we selected forecasts with lead times ranging from 1 to 24 hours initialized at 00 UTC and 12 UTC for our training set.

We selected 40 input variables at surface and pressure levels from the FuXi forecasts for downscaling task, guided by domain
90   knowledge (see Table 1). The dataset spans from May 2019 to October 2019 and July 2020 to August 2021, while the testing set includes data from November 2019 to June 2020 for testing the downscaling performance. The training set consists of 14,000 hours of data, whereas the testing set contains 6,346 hours, with no overlap between the two. Prior to inputting the data into the downscaling model, all input and output variables are normalized using z-score normalization.

Table 1. The details of input variables.

	Variable	Full Name	Unit
Press Level	U(50, 200, 500, 700, 850, 925, 1000, hPa)	U-component of wind	m s^{-1}
	V(50, 200, 500, 700, 850, 925, 1000, hPa)	V-component of wind	m s^{-1}
	Z(50, 200, 500, 700, 850, 925, 1000, hPa)	Geopotential	m^2 s^{-2}
	T(50, 200, 500, 700, 850, 925, 1000, hPa)	Temperature	K
	Q(50, 200, 500, 700, 850, 925, 1000, hPa)	Specific humidity	kg kg^{-1}
Surface	T$_{2M}$	2 meter temperature	K
	TP	Total precipitation	m h^{-1}
	U$_{10M}$	10 meter U-component of wind	m s^{-1}
	V$_{10M}$	10 meter V-component of wind	m s^{-1}
	MSL	Mean sea level pressure	Pa

2.2   Overview of Data Compression Framework

95   **Problem statement** We represent high-dimensional atmospheric data at a given time t using a tensor X^t with dimensions V $\times$ H $\times$ W, where H and W denote the number of latitude and longitude coordinates, respectively. For HRCLDAS, V=3, W=4384, H=6880. The indexing scheme $X^t_{v,h,w}$ indicates the value of variable v at time t and latitude-longitude coordinates (h, w).

The first objective of this study is to compress X^t to a lower-dimensional representation space Z^t using an encoder with learnable parameters within the proposed VAE compression framework. The second goal is to utilize the compressed data for
100   the downstream task of downscaling. Here, we select input Y^t_v, which represents a variable with a lead time of t from the FuXi forecast outputs. Our aim is to learn a mapping function f that converts the low-resolution input state Y_v at timestamp t to Z^t.

VAE data compression framework Fig. 1 illustrates the overall compression framework based on the VAE neural network. The VAE consists of an encoder E and a decoder D. The encoder E compresses the information from the current timestamp into a low-dimensional latent vector Z. The decoder reconstructs the original high-resolution data $X'_{t,i}$ from latent vector Z
105   through upsampling and convolution layers. To facilitate random sampling, E outputs a mean μ and standard deviation σ, followed by a sampling function that generates the latent representation Z.

Specifically, the encoder E first transforms the input to a higher-dimensional feature space using a $1{\times}1{\times}C$ convolutional layer, with C (the number of channels) initially set to 128. This is followed by four stages, where the first three each contain two ResNet blocks and a downsampling block. Each ResNet block utilizes two convolution layers with a kernel size of 3, followed by a Swish activation function (Ramachandran et al., 2017) and group normalization layers. Residual connections link the input and output, mitigating the risk of vanishing gradients.

The first convolutional layer of each stage doubles the feature map size, while the second convolutional layer maintains the same size. At the end of each stage, a downsampling block is applied, consisting of a convolutional layer with a kernel size of 3 and a stride of 2 to reduce the volume by half, replacing the traditional pooling operation. The fourth stage differs in that it does not include a downsampling layer. Instead, it uses two additional ResNet blocks to refine the compressed features. Finally, a convolutional layer with a feature map size of 4 is used to generate the mean μ and var σ, both with dimensions $548{\times}860{\times}4$.

The decoding process is a symmetric process with the encoding process. The decoder initially uses a convolutional layer to convert the channel dimension from 4 to 512. Bilinear interpolation method is used for upsampling at the end of each stage. Two ResNet blocks in the decoder further refine the features. The four decoding stages process the features sequentially, yielding feature map sizes of $548{\times}860{\times}4$, $548{\times}860{\times}512$, $548{\times}860{\times}512$, $1096{\times}1720{\times}256$, and $2192{\times}3440{\times}128$, respectively. Finally, a convolutional layer converts the channel dimension to the output dimension, resulting in an output size of $4384{\times}6880{\times}3$.

The loss component of the VAE consists of the reconstruction loss and the Kullback–Leibler divergence D_{kl}. We utilized the charbonnie loss (Lai et al., 2018) as reconstruction loss function as described in Eq. 1 to better handle the outliers situation. In addition, the Kullback–Leibler divergence D_{kl} regularizes the posterior distribution of the latent space $P = N(\mu, \sigma^2)$ to match the prior distribution, a standard Gaussian distribution $Q = N(0,1)$. Those parameters are optimized with the help of the re-parametrization trick (Kingma et al., 2016).

$$\mathcal{L}(X, X') = \sqrt{||X - X'||^2 + \epsilon^2} \tag{1}$$

$$KL(P||Q) = \sum_x P(X) log(\frac{P(X)}{Q(X)}) \tag{2}$$

Downscaling Model: For the downstream task, we selected downscaling as a case study to evaluate the usability of the compact data generated by our proposed VAE framework. The primary objective of the downscaling model is to enhance the output resolution of FuXi forecasts. Once the VAE model is trained, the HRCLDAS data are encoded by the VAE encoder to extract latent representations, which serve as ground truth for the downscaling model. We employed U-Net for this task (Ronneberger et al., 2015). The original FuXi forecast input has dimensions of $176{\times}276$ with a spatial resolution of 25 km. For downscaling, we interpolate the data to a resolution of $550{\times}3862{\times}3$ with a spatial resolution of 8 km.

U-Net utilizes a similar encoder-decoder architecture as the VAE. It comprises four stages, each containing two ResNet blocks in both the encoder and decoder. The decoder D_{unet}, which serves as an expansive path, employs upsampling layers to
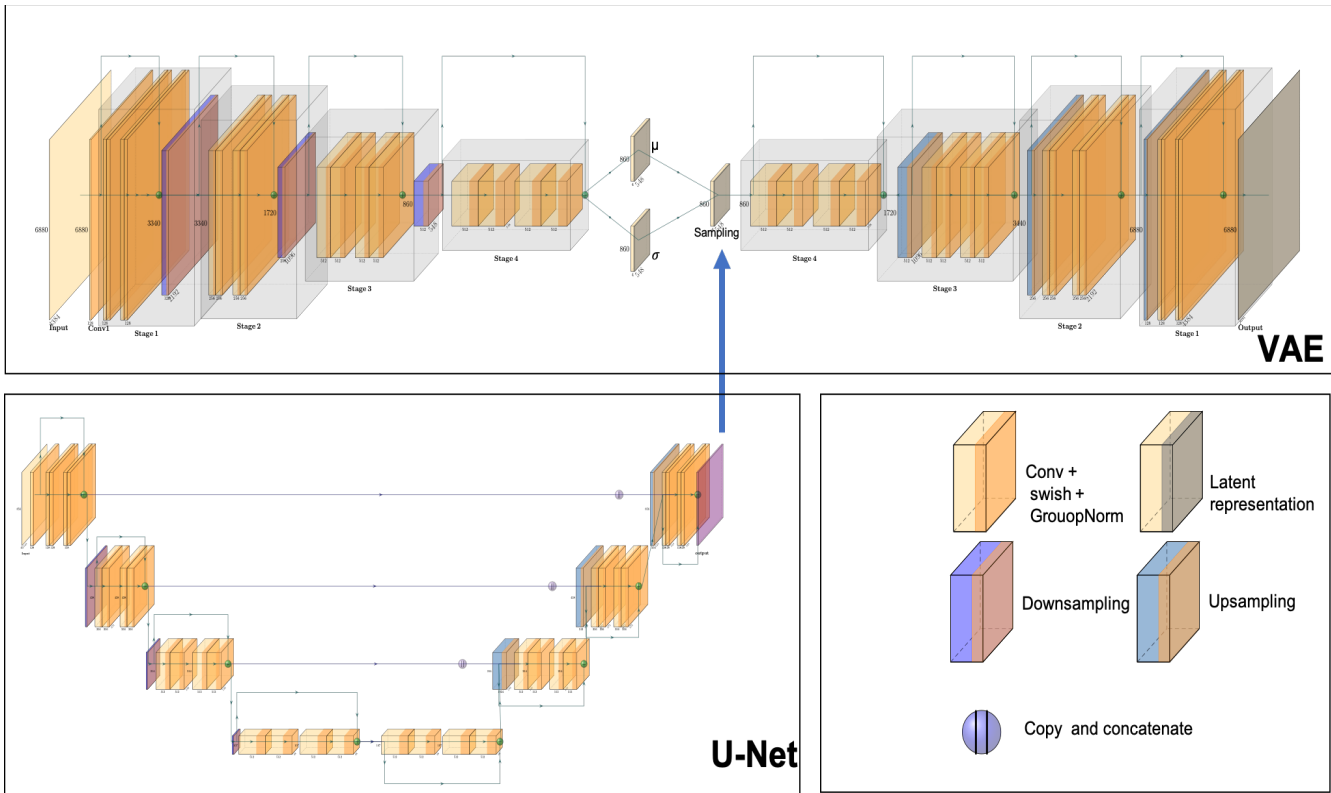
Figure 1. (Illustration of the proposed framework.

project the encoder output to the resolution of the compact data. The decoding part of U-Net is connected to the corresponding

140   encoder part via skip connections. A convolution operation is applied to the concatenated output of the up-sampled output and

its encoder counterpart, as illustrated in Fig. 1. Eventually, the U-Net trained on the original HRCLDAS data has 118 million

trainable parameters, while the U-Net trained on compressed data has 94.8 million parameters.

2.3   Implementation details

We first trained a VAE for data compression. The model was trained with a batch size of 8 and optimized using the Adam

145   optimizer with a learning rate of 1.6e-5. Given the high resolution of the original HRCLDAS data, we split it into 1000×1000

patches with overlaps in the latitude direction, resulting in 35 patches per dataset. To reduce computational costs and accelerate

convergence, we further divided these 1000×1000 patches into smaller 256×256 patches. We pre-trained the VAE model on

these smaller patches for the first 10 epochs and subsequently fine-tuned it using the larger patches for an additional 5 epochs.

Unlike natural images, where channels are highly correlated, weather variables exhibit weak inter-variable relationships. There-

150   fore, we trained separate VAE models for T_{2m}, and U_{10M}, V_{10M}, respectively.

For training the downscaling model, we used the compressed data Z^t with a resolution of 544×856 generated from the trained VAE encoder as ground truth. We trained the U-Net model with a batch size of 16 for 50 epochs. The model was optimized using the Adam optimizer with a learning rate of 3.2e-5. All the models were implemented using the PyTorch framework and trained on 8 NVIDIA A100 GPUs.

2.4 Experiment setup

Four experiments were conducted to assess the impact of different modeling strategies on data reconstruction performance within the proposed VAE framework. Additionally, three experiments were performed to evaluate the model's downscaling performance using the compact data generated by the VAE.

The four experiments within the VAE framework are detailed in Table 2. The baseline model (resize) employs a straightforward downsampling technique to reduce dimensions and an upsampling method for data reconstruction. To achieve improved reconstruction results, we trained the two VAE models using two different loss functions: $L1$ loss and Charbonnier loss, on 256 $\times$ 256 patches. Since the target variables do not exhibit linear correlations with one another, we trained separate models (VAE single variable) for each target variable, rather than treating them collectively, as is common with natural images. Finally, to obtain more accurate results, we fine-tuned the model trained on smaller patches using the full resolution data (VAE fine-tune).

For the downscaling task, we used bilinear interpolation (Inter) as the baseline for comparsion. We trained the U-Net model on both the original HRCLDAS data (No-VAE) and the compact data generated by the encoder from VAE (fine-tune) model, respectively, to access the impact of the data compression method on downstream task performance.

Table 2. The setting of different data compression methods

method	$L1$ loss	charbonnie loss 3	single variable	fine tune
Resize				
VAE ($L1$)	✓			
VAE(charbonnie loss)		✓		
VAE (single variable)		✓	✓	
VAE(fine-tune)		✓	✓	✓

2.5 Evaluation metrics

To evaluate the performance of the reconstruction and its downscaling capabilities, we use several metrics commonly applied in the weather and climate domain (Rasp et al., 2024; Zhong et al., 2024b; Gong et al., 2022). Specifically, we calculate the mean square error (MSE), root mean square error (RMSE), and power spectrum.

The MSE measures the difference between the ground truth $X_{w,h}$ and the reconstructed or downscaled results $X'_{w,h}$, where the grid corresponds to the cell center positions w and h in the zonal and meridional directions, respectively. RMSE, on the

other hand, represents the average difference between the ground truth and the reconstructed data generated by the model. Both
MSE and RMSE are negatively oriented, with an ideal value of 0 indicating perfect reconstruction.

$$MSE(v) = \frac{1}{N} \sum_{w=1}^{W} \sum_{h=1}^{H} \left[X'_{v,h,w} - X_{v,h,w} \right]^2. \tag{3}$$

$$RMSE(v) = \sqrt{\frac{1}{N} \sum_{w=1}^{W} \sum_{h=1}^{H} (X'_{v,h,w} - X_{v,h,w})^2} \tag{4}$$

In addition, to evaluate the local-scale variability and determine the information preserved or lost by the VAE framework for
the downstream tasks, we conduct power spectrum analysis following the methodology described by Rasp et al. (2024). The
power spectrum is calculated along lines of constant latitude as a function of wavenumber (unitless), frequency (km−1), and
wavelength (km). The Discrete Fourier Transform (DFT), denoted as F_K is computed using the following equation:

$$F(k) = \frac{1}{L} \sum_{l=0}^{L-1} f_l e^{\frac{-i2\pi kl}{L}} \tag{5}$$

Since our evaluation focuses on the regional rather than global scales, we do not consider the circumference in this study.
The power spectrum for constant latitude is obtained using the following equation:

$$Sk = 2|F_k|^2, k = 1, 2, ..., L/2 \tag{6}$$

Additionally, we approximate the average zonal power spectrum using:

$$\int_0^l |fl| dl \approx \sum_{k=0}^{L/2} S_k \tag{7}$$

In addition to the meteorological evaluation, we incorporate the structural similarity index (SSIM) (Wang et al., 2004), which
was initially developed for computer vision to quantify and compare spatial variability between downscaled fields and ground
truth data. This metric has recently been applied in meteorological research (Zhong et al., 2024b; Gong et al., 2022).

$$SSIM(y, \hat{y}) = \frac{(2\mu_{\hat{y}}\mu_y + C_1) + (2\sigma_{\hat{y}y} + C_2)}{(\mu_{\hat{y}^2} + \mu_{y^2} + C_1)(\sigma_{\hat{y}^2)} + \sigma_{y^2} + C_2)} \tag{8}$$

Table 3. The performance in terms of RMSE for different data compression models, where U_{10M}, V_{10M}, and T_{2M} represent 10 meter U-component of wind, 10 meter V-component of wind, and 2 meter temperature.

RMSE	U_{10M}	V_{10M}	T_{2M}
Resize	0.0771	0.0783	0.7109
VAE ($L1$)	0.108	0.100	0.752
VAE(charbonnie loss)	0.104	0.090	0.578
VAE (single variable)	0.0428	0.029	0.353
VAE(fine-tune)	**0.0181**	**0.0139**	**0.124**

3 Experiments results

3.1 Representation performance

Table 3 presents performance results for various data compression strategies, including simple resize, VAE with $L1$ loss, VAE with Charbonnier loss, VAEs trained on single variables, and the VAE fine-tuning strategy. Among these methods, the VAE fine-tuning strategy demonstrates superior performance in reconstructing U_{10M}, V_{10M}, and T_{2M}. Additionally, we observed that simple VAE with $L1$ and Charbonnie loss struggles to effectively represent the temperature and wind fields, leading to a high RMSE. Notably, training on a single variable is consistently superior to training on multiple variables. Furthermore, a comparison between the VAE trained from scratch (single variable) and the VAE fine-tuned from pretraining with small patches to full resolution indicates that the pretraining method significantly enhances the representation performance. The VAE fine-tuning approach achieves RMSE values of 0.0181 K, 0.0139 K, and 0.124 K for the three variables respectively, which is more than 4 times better than the linear interpretation method.

To further understand the VAE's performance in data compression, particularly for extreme values, we examine the differences in density distributions among the ground truth, VAE, and a simple resize baseline model, as illustrated in Fig. 3. For T_{2M} (Fig. 3a), the VAE-reconstructed data closely aligns with the frequency of positive temperature values in the ground truth HRCLDAS data. In contrast, the resize method results in a slight decrease in the frequency of positive temperatures. For negative temperatures, the VAE model exhibits a lower frequency compared to the original HRCLDAS data. Regarding U_{10M} and V_{10M} (Fig. 3b and Fig. 3c), the VAE effectively preserves a higher frequency of both extreme high and low values.

3.2 Downscaling performance

In our study, we selected the fine-tuned VAE (fine-tune) as the data compression method, subsequently applying the compressed data produced by VAE (fine-tune) for our downstream task - downscaling. Figure 4 and Figure 5 demonstrate the performance of the UNet model trained on compressed HCLDAS data (VAE) and original data (No-VAE) in terms of MSE (Fig. 4) and SSIM(Fig. 5) respectively, compared to the baseline model - interpolation method, across lead times from 1 to 18 hours. The results indicate that the deep learning-based model significantly outperforms the simple interpolation method (Inter). Specifically, the mean square error values for linear interpolation, No-VAE, and VAE are 11.58 K^2, 4.965,79K^2, and
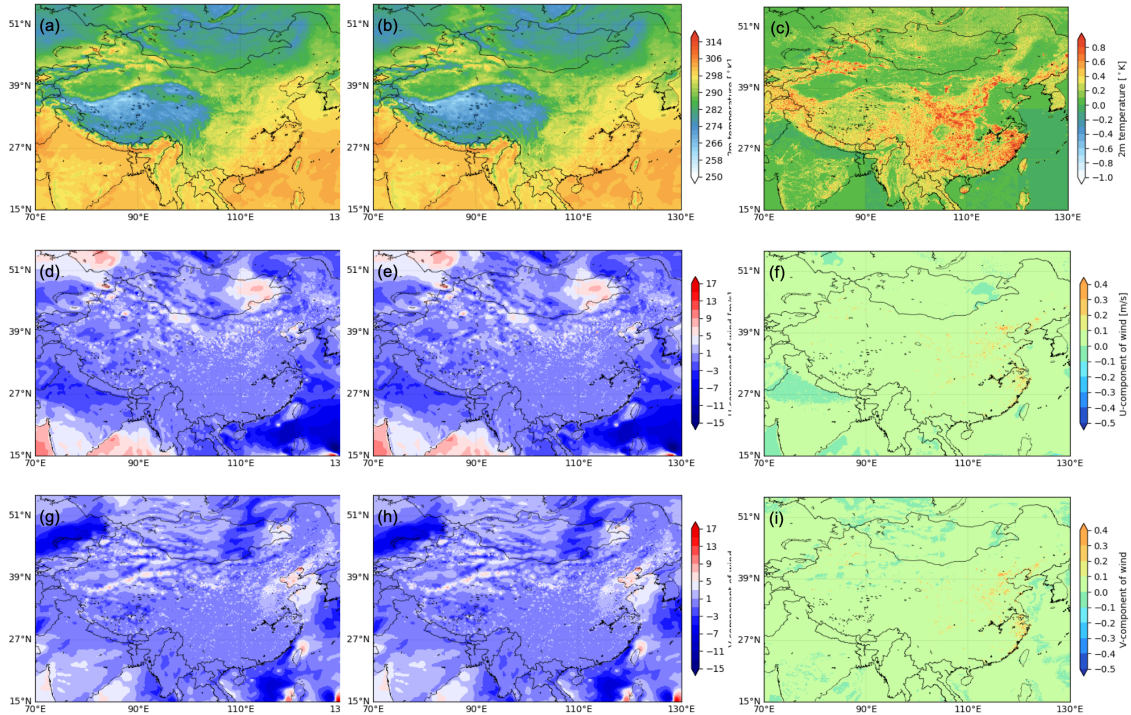
Figure 2. Example reconstructed fields by VAE (fine-tune) for T_{2M}, U_{10M} and V_{10M} respectively on 9th September 2021, 00 UTC. (a) T_{2M} from HRCLDAS, (b) the reconstructed T_{2M} field by VAEs (fine-tune), (c) differences bet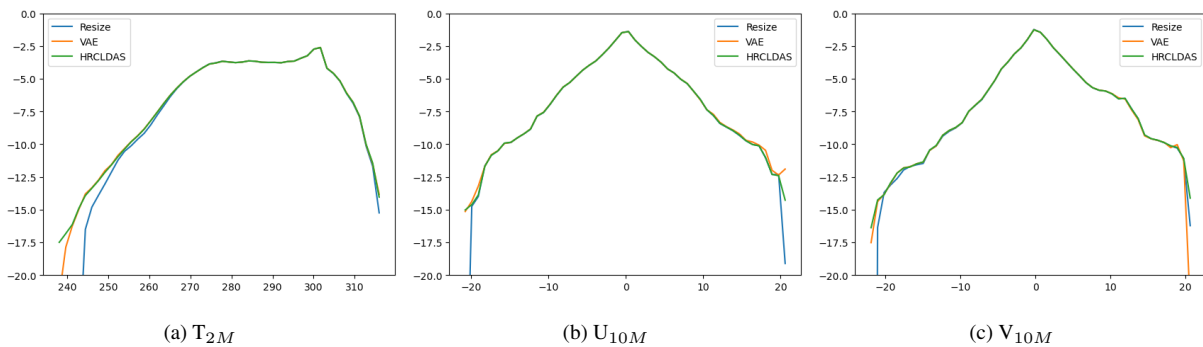ween HRCLDA T_{2M} and reconstructed T_{2M}, (d) U_{10M} from HRCLDAS, (e) the reconstructed U_{10M} field by VAEs (fine-tune), (f) differences between HRCLDA U_{10M} and reconstructed U_{10M}, (g) V_{10M} from HRCLDAS, (h) the reconstructed V_{10M} field by VAEs (fine-tune), (g) differences between HRCLDA V_{10M} and reconstructed V_{10M}.

| (a) T_{2M} | (b) U_{10M} | (c) V_{10M} |

Figure 3. Displaying of log(density) plot for (a) T_{2M}, (b) U_{10M}, and (c) V_{10M} by the VAE method comparing to the resize method and HRCLDAS data.

$5,79K^2$, respectively, demonstrating that the No-VAE and VAE outperform linear interpolation by approximately 53%, 60% for temperature in terms of MSE.

10

220   We also employed the SSIM, commonly used in downscaling tasks to access the perceptual similarity between images. SSIM quantifies and compares mean and global spatial variability in the downscaled fields against the ground truth and also accounts for covariances. The SSIM comparisons for all models are presented in Fig 5, showing that both VAE and No-VAE consistently outperform the interpolation method across all lead times and variables.

        Furthermore, Figure 6 presents the average zonal power spectrum plots for the No-VAE and VAE downscaling models
225   at a 1-hour lead time, compared to the ground truth HRCLDAS data and the baseline interpolation method. The findings reveal that the baseline interpolation method fails to capture significant details across all scales for the three variables. This is expected, as interpolation tends to average out variances, leading to a smoother appearance, as visually confirmed in Fig 8 and Fig. 10. In contrast, both the VAE and No-VAE models preserve more details, with their performances being quite similar. This suggests that the compact data representation from the VAE does not lead to substantial information loss in our
230   proof-of-concept downscaling task compared to using the original HRCLDAS data.

        It is worth mentioning that the No-VAE model generates artifacts at small scales for T_{2M} as the examples demonstrated in Fig. 8c and Fig. 8f. Sudden temperature changes create strong contrasts in specific areas, such as northeast China. This issue arises from the unpatching process, where small patches are reconstructed into the original image due to computational limitations. In contrast, this problem can be effectively mitigated by feeding the U-Net the entire latent features of HRCLDAS
235   data directly from the VAE model, resulting in a smoother downscaling field, as demonstrated in Fig. 8d and Fig. 8g.

        In addition, we observed that both the VAE and No-VAE models struggle to capture fine-scale features of T_{2M}. This limitation is likely due to the significant elevation variations in the Himalayas and certain regions of China, where topography is crucial for maintaining local temperature gradients. Since the primary focus of this study is to validate the VAE compression method, we did not incorporate topography data in the current models. However, this will be considered in future research.
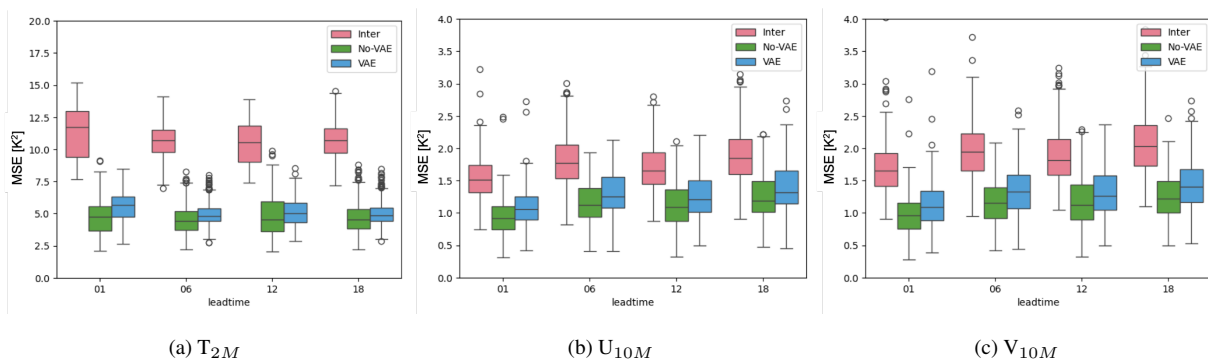

(a) T_{2M} (b) U_{10M} (c) V_{10M}

Figure 4. Box-and-whisker plot for comparision in terms of MSE between baseline and machine learning methods for (a) T_{2M}, (b) U_{10M}, and (c) V_{10M} with lead time of 1h to 18h. The solid horizon indicates the minimum and maximum MSE, excluding outliers; the box bounds the interquartile range from the 25th to 75th percentiles, with the 50th percentile. The red color indicates the baseline model "resize"; the green color indicates the U-Net trained on original HRCLDAS data; the blue color indicates the U-Net trained on compact HRCLDAS data generated by VAE.
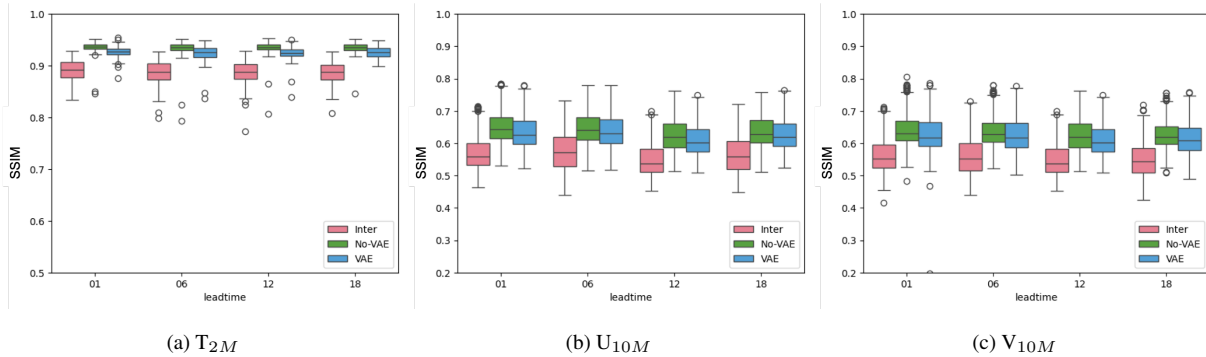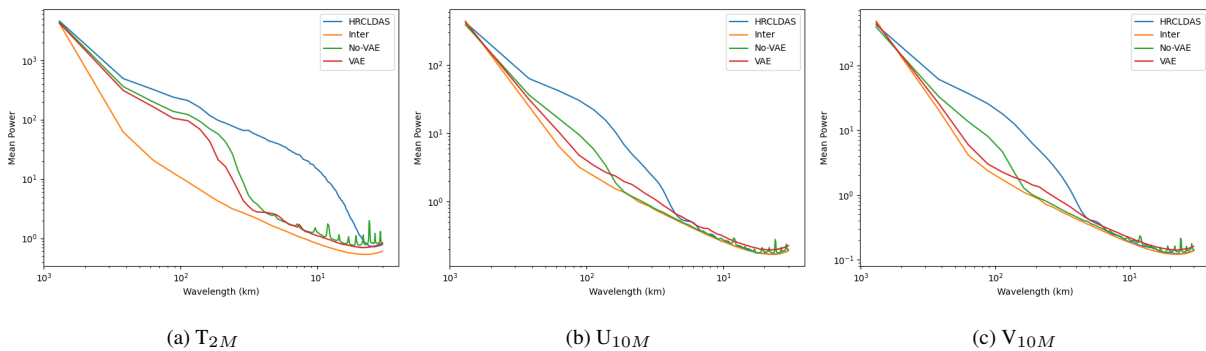
(a) T_{2M} (b) U_{10M} (c) V_{10M}

Figure 5. The Box-and-whisker plot for comparsion in terms of SSIM between baseline and machine learning methods for (a) T_{2M}, (b) U_{10M}, and (c) V_{10M} with a lead time of 1 h to 18 h. The solid horizon indicates the minimum and maximum MSE, excluding outliers; the box bounds the interquartile range from the 25th to 75th percentiles, with the 50th percentile. The red color indicates the baseline model "resize"; the green color indicates the U-Net trained on original HRCLDAS data; the blue color indicates the U-Net trained on compact HRCLDAS data generated by VAE.

(a) T_{2M} (b) U_{10M} (c) V_{10M}

Figure 6. Plot displaying of the power spectrum for (a) T_{2M}, (b) U_{10M}, and (c) V_{10M} with a lead time of 1 h by the VAE method comparing to the interpolation, Non-VAE, and HRCLDAS ground truth data. The x-axis indicates the number of wavelength (km) and y-axis is the number average power spectrum.

To better understand the realism of downscaling performance, we selected three examples to compare the performance of temperature and zonal wind variables using "eyeball" analysis (see Fig 8 and Fig 10). These examples correspond to 1st November 2019, at 01:00 UTC, 15th April 2020, at 18:00 UTC, and 15th January 2020, at 12:00 UTC, representing FuXi forecasts with 1-hour, 6-hour, and 12-hour lead times, respectively. As illustrated in Fig 8, it can be seen that the interpretation method shows significant biases in temperature downscaling for southwestern China, particularly within the complex mountainous terrain of the Himalayas (see topography map in Fig .7). This bias was especially pronounced in northeastern China during the cold winter period on 15th January 2020, when the No-VAE interpolation methods were applied. In contrast, the application of the VAE method with the compressed data allowed the U-Net to more effectively capture and reduce errors in this region.
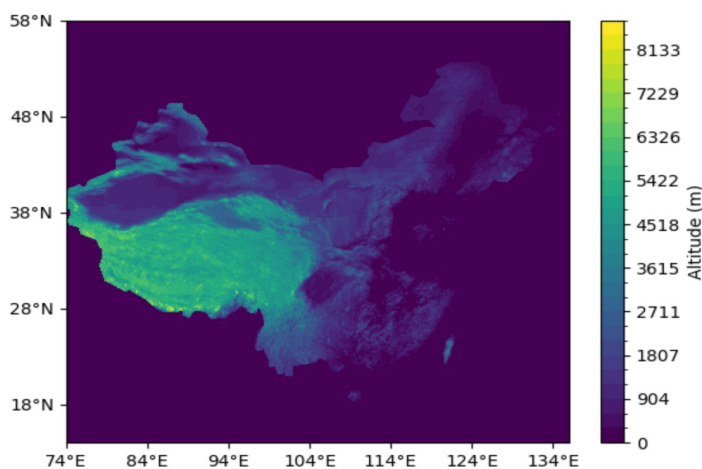
Figure 7. Topography height of the surface in the target region.

4 Conclusion and Discussion

This study investigates the compression of high-dimensional, high-volume weather data using the NIC method, inspired by
250   advancements in the computer vision domain, for downstream weather and climate applications. Specifically, we employed
our proposed VAE framework to compress three years of high-resolution HRCLDAS data, reducing the data size from 8.61
TB to a compact 204 GB, an impressive 42-fold reduction in storage requirements. The results demonstrate that the VAE
framework, enhanced with a fine-tuning strategy, outperforms other baseline methods, significantly reducing reconstruction
errors, thereby providing more accurate reconstructions of 2-meter temperature, 10-meter U-component of wind, and 10-meter
255   V-component of wind, in terms of RMSE. Moreover, the VAE-based model demonstrates superior performance in preserving
extreme values, effectively maintaining both high and low values from the original HRCLDAS dataset.

To further validate the effectiveness and usability of the compact data generated by the VAE framework, we used it for
the downscaling task. The performance metrics reveal no significant differences in MSE and SSIM between the original
HRCLDAS data and compact HRCLDAS data. Although a slight degradation in performance was observed with the com-
260   pact data, likely due to the reduced number of parameters utilized by the U-Net model. Further evaluation based on the power
spectrum suggests that both the original HRCLDAS and compact data significantly outperform the baseline bilinear interpo-
lation method. While using the compact data results in a minor loss of information at larger scales, it effectively retains key
information at finer scales for all three target variables, similar to the original HRCLDAS data. This demonstrates the usability
and effectiveness of the compact data for downstream downscaling tasks. Looking ahead, this approach has the potential to be
265   expanded to other applications, such as large-scale weather and climate forecasting, where the need for processing extensive
datasets is more crucial.

Code and data availability. The source code used for training and running VAE compression and downscaling models in this work is available at https://doi.org/10.5281/zenodo.13862076.

Author contributions. QL BG, and XRZ equally contributed to this work.

270 *Competing interests.* The authors declare that they have no conflict of interest

References

Ballé, J., Laparra, V., and Simoncelli, E. P.: End-to-end optimized image compression, arXiv preprint arXiv:1611.01704, 2016.

Ballé, J., Minnen, D., Singh, S., Hwang, S. J., and Johnston, N.: Variational image compression with a scale hyperprior, arXiv preprint arXiv:1802.01436, 2018.

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., and Tian, Q.: Accurate medium-range global weather forecasting with 3D neural networks, Nature, 619, 533–538, 2023.

Chen, L., Zhong, X., Zhang, F., Cheng, Y., Xu, Y., Qi, Y., and Li, H.: FuXi: a cascade machine learning forecasting system for 15-day global weather forecast, npj Climate and Atmospheric Science, 6, 190, 2023.

Chen, Z., Lu, G., Hu, Z., Liu, S., Jiang, W., and Xu, D.: LSVC: A learning-based stereo video compression framework, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6073–6082, 2022.

Cheng, Z., Sun, H., Takeuchi, M., and Katto, J.: Learned image compression with discretized gaussian mixture likelihoods and attention modules, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 7939–7948, 2020.

Gong, B., Langguth, M., Ji, Y., Mozaffari, A., Stadtler, S., Mache, K., and Schultz, M. G.: Temperature forecasting by deep learning methods, Geoscientific model development, 15, 8931–8956, 2022.

Han, S., Liu, B., Shi, C., Liu, Y., Qiu, M., and Sun, S.: Evaluation of CLDAS and GLDAS Datasets for Near-Surface Air Temperature over Major Land Areas of China, Sustainability, 12, https://doi.org/10.3390/su12104311, 2020a.

Han, S., Liu, B., Shi, C., Liu, Y., Qiu, M., and Sun, S.: Evaluation of CLDAS and GLDAS datasets for Near-surface Air Temperature over major land areas of China, Sustainability, 12, 4311, 2020b.

Harris, L., McRae, A. T., Chantry, M., Dueben, P. D., and Palmer, T. N.: A Generative Deep Learning Approach to Stochastic Downscaling of Precipitation Forecasts, arXiv preprint arXiv:2204.02028, 2022.

Hersbach, H. et al.: The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146, 1999–2049, 2020.

Kingma, D. P., Salimans, T., Jozefowicz, R., Chen, X., Sutskever, I., and Welling, M.: Improved variational inference with inverse autoregressive flow, Advances in neural information processing systems, 29, 2016.

Klöwer, M., Razinger, M., Dominguez, J. J., Düben, P. D., and Palmer, T. N.: Compressing atmospheric data into its real information content, Nature Computational Science, 1, 713–724, 2021.

Lai, W.-S., Huang, J.-B., Ahuja, N., and Yang, M.-H.: Fast and accurate image super-resolution with deep laplacian pyramid networks, IEEE transactions on pattern analysis and machine intelligence, 41, 2599–2613, 2018.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W., et al.: GraphCast: Learning skillful medium-range global weather forecasting, arXiv preprint arXiv:2212.12794, 2022.

Leinonen, J., Nerini, D., and Berne, A.: Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network, IEEE Transactions on Geoscience and Remote Sensing, 59, 7211–7223, 2020.

Li, H., Chen, L., Zhong, X., Zhang, F., Cheng, Y., Xu, Y., and Qi, Y.: FuXi: A cascade machine learning forecasting system for 15-day global weather forecast, 2023.

Ramachandran, P., Zoph, B., and Le, Q. V.: Searching for activation functions, arXiv preprint arXiv:1710.05941, 2017.

Rasp, S., Hoyer, S., Merose, A., Langmore, I., Battaglia, P., Russell, T., Sanchez-Gonzalez, A., Yang, V., Carver, R., Agrawal, S., et al.: WeatherBench 2: A benchmark for the next generation of data-driven global weather models, Journal of Advances in Modeling Earth Systems, 16, e2023MS004 019, 2024.

Ronneberger, O., Fischer, P., and Brox, T.: U-net: Convolutional networks for biomedical image segmentation, in: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, pp. 234–241, Springer, 2015.

315 Shi, C., Jiang, L., Zhang, T., Xu, B., and Han, S.: Status and Plans of CMA Land Data Assimilation System (CLDAS) Project, in: EGU General Assembly Conference Abstracts, EGU General Assembly Conference Abstracts, p. 5671, 2014.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P.: Image quality assessment: from error visibility to structural similarity, IEEE transactions on image processing, 13, 600–612, 2004.

Zhong, X., Chen, L., Fan, X., Qian, W., Liu, J., and Li, H.: FuXi-2.0: Advancing machine learning weather forecasting model for practical
320 applications, https://arxiv.org/abs/2409.07188, 2024a.

Zhong, X., Du, F., Chen, L., Wang, Z., and Li, H.: Investigating transformer-based models for spatial downscaling and correcting biases of near-surface temperature and wind-speed forecasts, Quarterly Journal of the Royal Meteorological Society, 150, 275–289, 2024b.
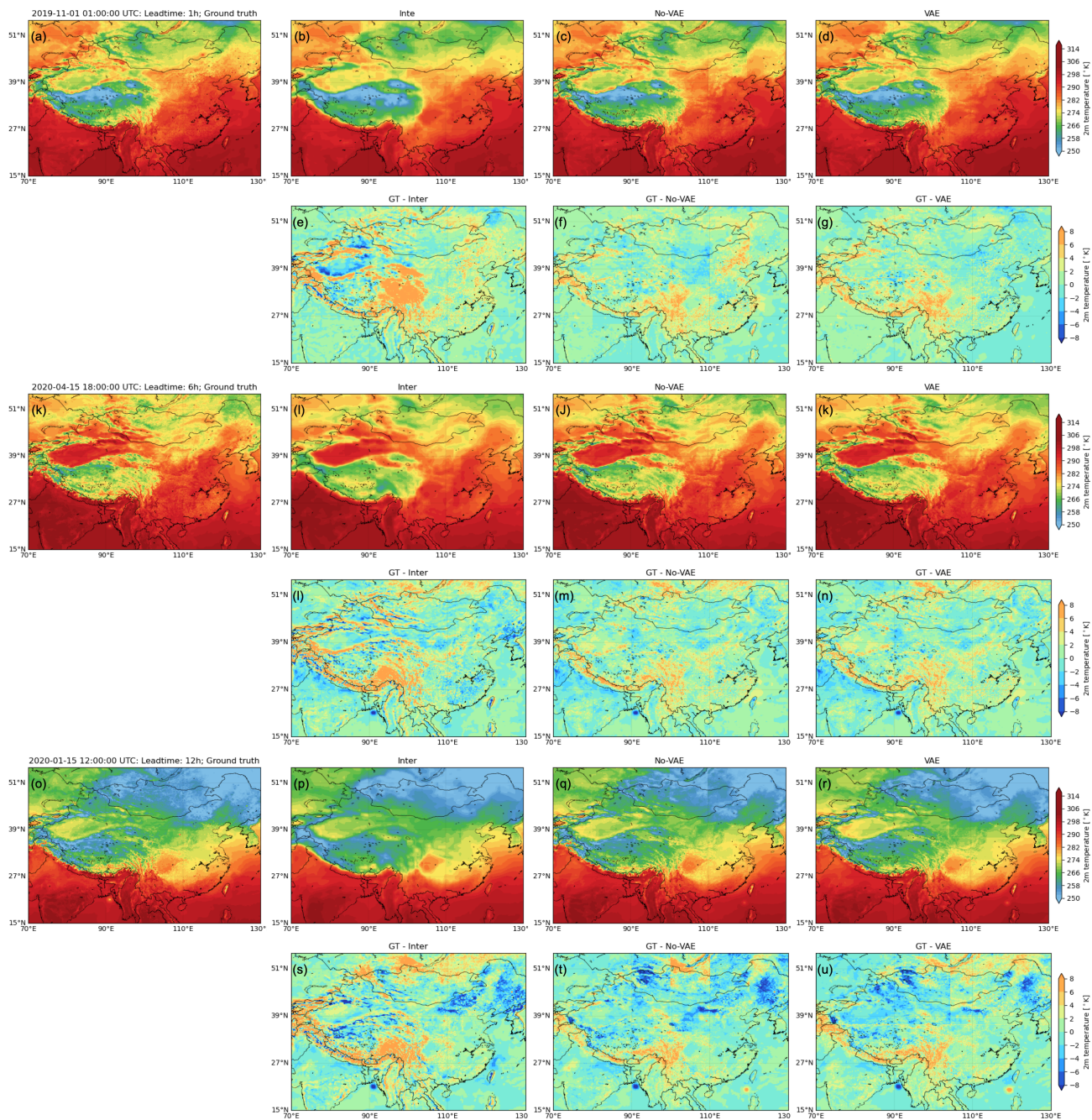
Figure 8. Example downscaling for 2 m temperature with a lead time of 1 h (panels: b - d), 6 h (panels: i-k) and 12 h (p-f), by interpretation, No-VAE, and VAE methods respectively. (a,h,o) are the ground truth from the HRCLAS dataset. (e-g, l-n, s-u) The difference between the ground truth and the downscaled fields generated by interpretation, No-VAE, and VAE method at the time of 1st November 2019, 01:00 UTC, 15th April 2020, 18:00 UTC, and 15th December 2020, 12:00 UTC.
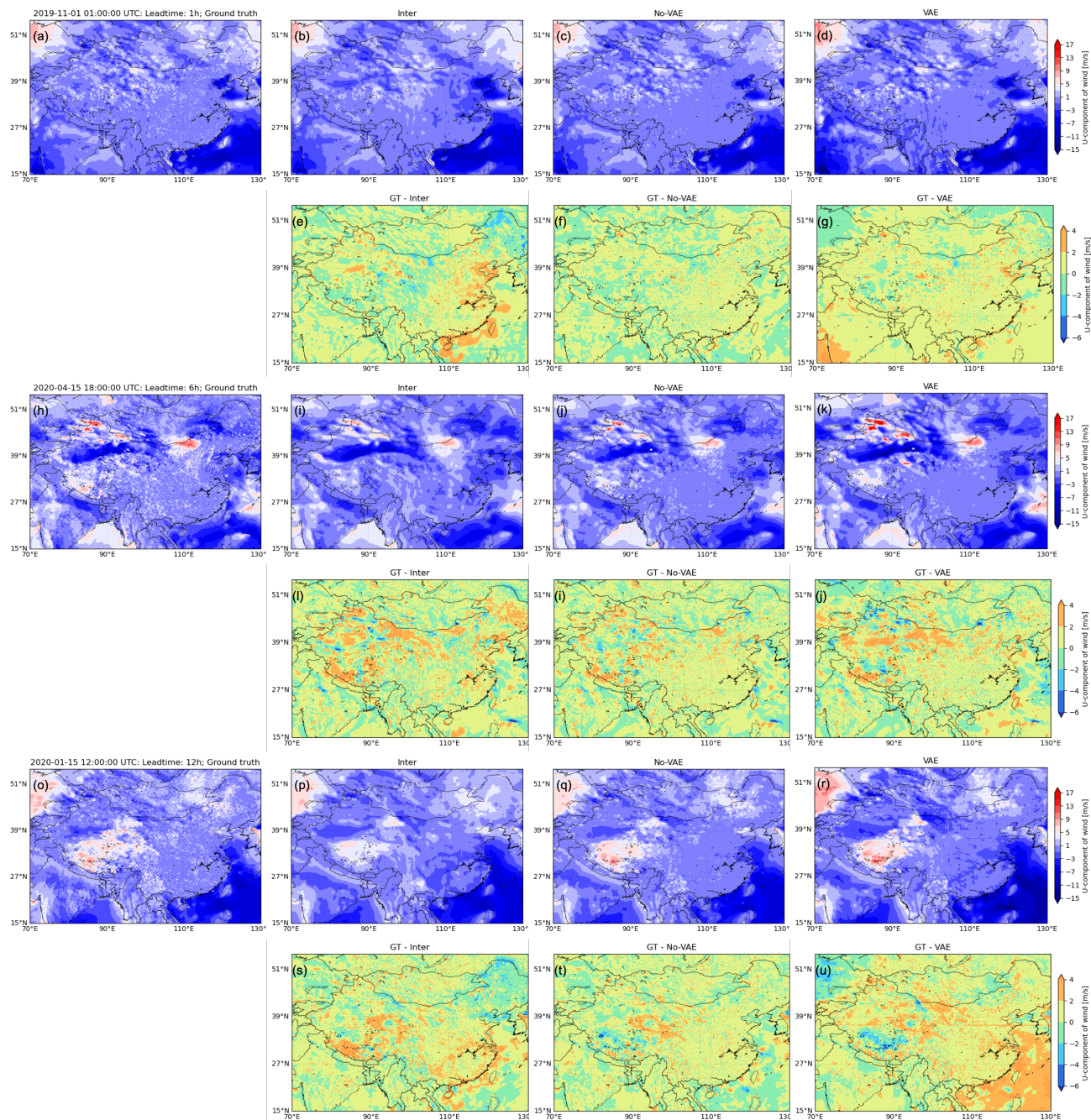
Figure 9. Example downscaling for U-component of wind with a lead time of 1 h (panels: b - d), 6 h (panels: i-k) and 12 h (p-f), by interpretation, No-VAE, and VAE methods respectively. (a,h,o) are the ground truth from the HRCLAS dataset. (e-g, l-n, s-u) The difference between the ground truth and the downscaled fields generated by interpretation, No-VAE, and VAE method at the time of 1st November 2019, 01:00 UTC, 15th April 2020, 18:00 UTC, and 15th December 2020, 12:00 UTC.
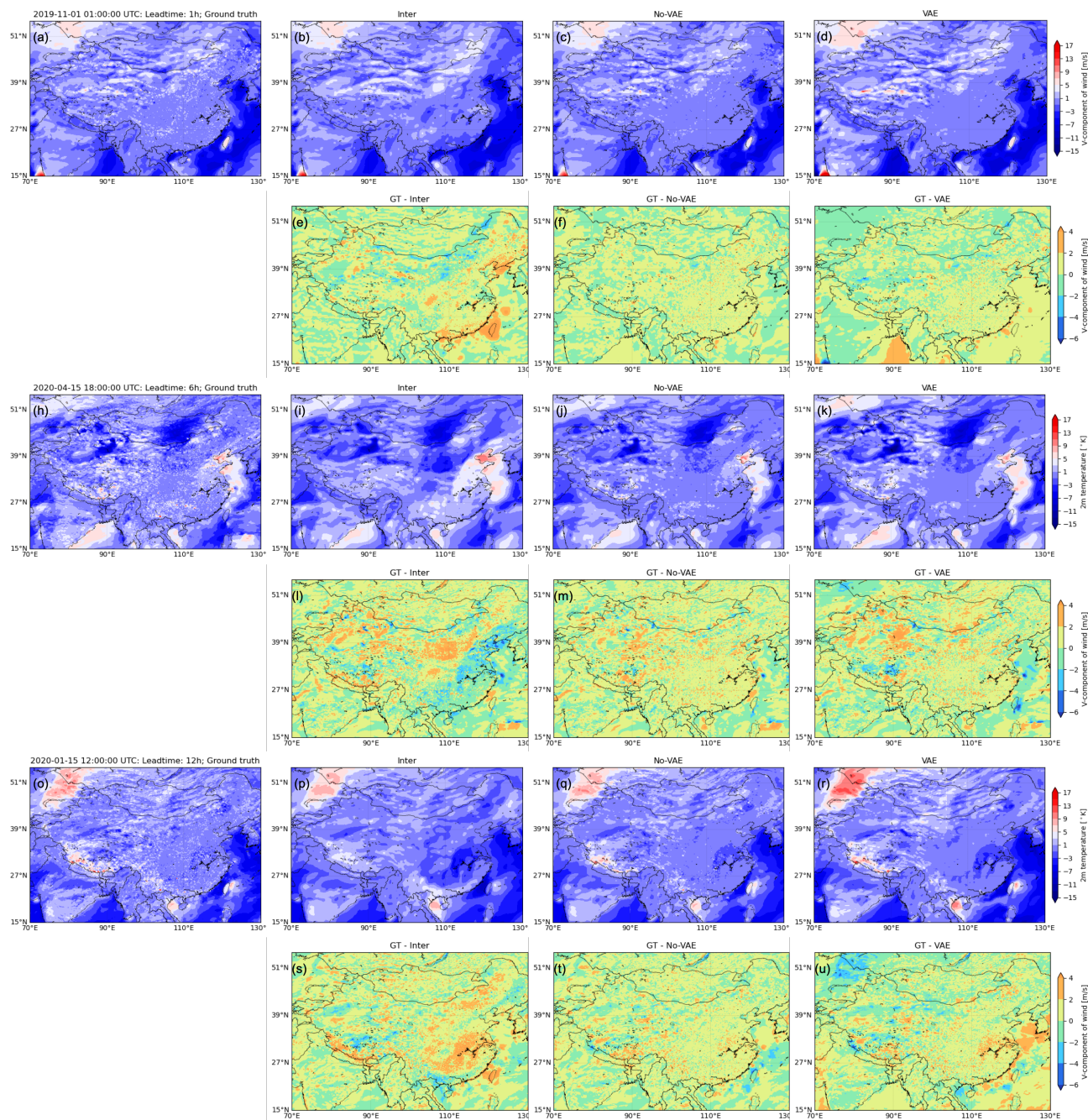
Figure 10. Three examples of downscaling for 10 meter V-component of wind with a lead time of 1 h (panels: b - d), 6 h (panels: i-k) and 12 h (p-f), by interpretation, No-VAE, and VAE methods respectively. (a,h,o) are the ground truth from the HRCLAS dataset. (e-g, l-n, s-u) The difference between the ground truth and the downscaled fields generated by interpretation, No-VAE, and VAE method at the time of 1st November 2019, 01:00 UTC, 15th April 2020, 18:00 UTC, and 15th December 2020, 12:00 UTC.